CLAIMS

I Claim:

1. A method for identifying a computer virus in interpreted language source code, the method comprising:

receiving a portion of interpreted language source code;

generating a language-independent representation of the portion of the interpreted language source code;

comparing the language-independent representation with a virus signature; and

determining if the language-independent representation matches the virus signature, whereby a match indicates a computer virus has been identified.

2. The method of claim 1, wherein the interpreted language source code is a scripting language source code.

3. The method of claim 1, wherein the virus signature is a language-independent representation of an interpreted language source code computer virus.

4. The method of claim 1, wherein the portion of interpreted language source code and the virus signature are represented as a linearized string of key actions.

5. A method for generating a virus signature, the method comprising:

receiving a portion of interpreted language source code containing a computer virus;

generating a language-independent representation of the computer virus; and

storing the language-independent representation of the computer virus as a virus signature.

6. The method of claim 5, wherein the interpreted language source code is a scripting language source code.

7. The method of claim 5, wherein the virus signature is compiled in binary format.

8. The method of claim 5, wherein the language independent representation is a linearized string of key actions.

9. The method of claim 5, wherein the virus signature includes input from a virus analyst.

10. The method of claim 5, further comprising:

parsing the portion of interpreted language source code into tokens; and

generating the language-independent representation of the computer virus using at least a portion of the tokens.

11. A method for identifying a virus in interpreted language source code, the method comprising:

receiving a portion of interpreted language source code;

parsing the portion of the interpreted language source code into tokens to generate a tokenized source code, wherein at least some of the tokens represent key actions;

extracting selected key actions from the tokenized source code,

linearizing the key actions to generate an executing thread;

comparing the executing thread with a virus signature of a known virus; and

determining whether the executing thread matches the virus signature.

12. The method of claim 11, further comprising:

outputting the identification of the known virus.

13. The method of claim 11, wherein the portion of the interpreted language source code is lexically parsed.

14. The method of claim 11, wherein the portion of the interpreted language source code is lexically and grammatically parsed.

15. A method for generating a virus signature from a portion of interpreted language source code including a computer virus, the method comprising:

receiving a portion of interpreted language source code containing a computer virus;

parsing the portion of the interpreted language source code containing the computer virus into tokens to generate tokenized source code, wherein at least some of the tokens represent key actions;

extracting key actions from the tokenized source code,

linearizing the key actions to generate an executing thread;

determining the set of minimum key actions in the executing thread required to effect the computer virus; and

storing the set of minimum key actions as a virus signature.

16. The method of claim 15, further comprising:

compiling the virus signature in binary format.

17. The method of claim 15, further comprising:

compiling the virus signature with data input by a virus analyst; and

storing the virus signature as part of a virus pattern file.

18. The method of claim 17, wherein the virus pattern file further includes a dictionary of key actions.

19. The method of claim 15, wherein the portion of the interpreted language source code is lexically parsed.

20. The method of claim 15, wherein the portion of the interpreted language source code is lexically and grammatically parsed.

21. A computer readable medium containing program code for identifying a computer virus in interpreted language source code, the computer readable medium comprising instructions for:

receiving a portion of interpreted language source code;

parsing the portion of the interpreted language source code into tokens to generate a tokenized source code, wherein at least some of the tokens represent key actions;

linearizing at least a portion of the key actions to generate an executing thread;

comparing the executing thread with a virus signature of a known computer virus; and

determining whether the executing thread matches the virus signature.

22. The computer readable medium of claim 21, further comprising:

outputting the identification of the known computer virus.

23. The computer readable medium of claim 21, wherein the portion of the interpreted language source code is lexically parsed.

24. The computer readable medium of claim 21, wherein the portion of the interpreted language source code is lexically and grammatically parsed.

25. A computer readable medium containing program code for generating a virus signature from a portion of interpreted language source code including a computer virus, the computer readable medium comprising instructions for:

receiving a portion of interpreted language source code containing a computer virus;

parsing the portion of the interpreted language source code containing the computer virus into tokens to generate tokenized source code, wherein at least some of the tokens represent key actions;

linearizing at least a portion of the key actions to generate an executing thread;

determining the set of minimum key actions in the executing thread required to effect the computer virus; and

storing the set of minimum key actions as a virus signature.

26. The computer readable medium of claim 25, further comprising:

compiling the virus signature in binary format.

27. The computer readable medium of claim 25, further comprising:
compiling the virus signature with data input by a virus analyst; and
storing the virus signature as part of a virus pattern file.

28. The computer readable medium of claim 27, wherein the virus pattern file further includes a dictionary of key actions.

29. The computer readable medium of claim 25, wherein the portion of the interpreted language source code is lexically parsed.

30. The computer readable medium of claim 25, wherein the portion of the interpreted language source code is lexically and grammatically parsed.